

# Weekly Report

Apr. 8<sup>th</sup>, 2018

## Done:

1. 参加一个 user study.
2. 看论文及一些思考.

I looked at the website of the VISREG workshop. Here is a summary and my thoughts.

(1) The two organizations' focus are medical data (volume data, SciVis). Some of them can be viewed as visual summary, but requires domain knowledge to understand.

(2) [Technique scope page](#) elaborate what visual summary do -- help to better understand the different sources that need to be analyzed, and can be used to describe the variability of the data. It provide a **useful starting point** for an exploratory analysis in common overview-first, details-on-demand setting.

Visual report is one form of visual summary: it can **communicate** the complexity and variability of a dataset (**findings in the data**) to other people with a fixed format, and no interaction allowed. This restrictions make more generating a report challenging.

(3) [Call for paper](#) page lists 3 research directions and topics, and corresponding example papers: **Data summarization, report generation, and visualization techniques for reports and summaries.**

Among the examples of data summarization, an EuroVis18 paper: (attached 1-design factors for summary vis) surveys the **summarization methods, purpose, tasks, and data type** of summary visualization. Many of the visual summaries here are extracted from visual analysis systems (the visual summaries usually serve as a starting point of analysis/roadmap to guide user interactions). Some summarization methods support certain tasks.

(4) [Schedule page](#) present the accepted papers but only one is available online. It's a data quality report card, designed based on user requirements. They also design a card browser to support simple actions on choosing the report card.

## (5) My reflection:

(about how to define visual fusion and data fusion, as mentioned in this [doc](#)) We can borrow something from the survey paper in (3), for example, the **data summary methods** and **task taxonomy**. However, they target at visual summary, and visual report is still different.

a. We can extend the data summary methods, transform into data fusion.

note: The survey summarize the summary methods as: Aggregation, Subsampling, Filtering and Projection. However, take wildfire data as an example, we can compute some values

like: correlation between different weather time series, abrupt changes of fire size or wind speed...so we can add another item here.

b. Those tasks can't be fulfilled with a report (because of no interaction allowed) should be removed from task taxonomy. Note that, tasks can be fulfilled by either data fusion or visual fusion. So, →

c. We have to define visual fusion. In fact, I think the juxtaposition, superimposition, overloading, nesting, and integration (see the 2nd attached paper Composite Visualization) is enough. Even annotation can be deemed as juxtaposition of visual elements.

d. Finally, two questions:

Q1: the wildfire data/scenarios doesn't include all those data or visual fusion steps. So probably we need some more scenarios.

Q2: Second, suppose I have several time-series. (a) I can simply visualize them as multiple line charts and stack them; or, (b) I can first calculate the variations of them and correlation coefficient between them, then encode these values beside the stacked line charts.

The problem is, for (a), obviously it's visual fusion; For (b), I first do data fusion, then when I encoding them beside the line charts, it can also be one type of visual fusion. How can we discriminate between them?

#### **To Do:**

1. 归纳两种 fusion 的特点, 主要方法是阅读一些实际的 application 和 design study. 切入是, 如果是 report, summary 相关的最好. 在下面提到的第一篇综述已有不少 visual summary 相关.
2. 准备各种材料, 去旧金山德国使馆面签. 考虑来回路上时间可能需要一天.

#### **Paper Reading:**

1. EuroVis18 *Design Factor for Summary Visualization in Visual Analytics* 关于 summary visualization 的一个归纳. 首先放下 summary vis 的定义: “the result of an explicit set of design decisions that compress and/or simplify data for display, which includes choices of data reduction methods and visual representations.” 大部分列举的文章都是一个系统的主视图或者某个 design study 的核心部件. 这里的 factor 分成四个部分, 分为 data summarization methods, purpose, analysis tasks, and data. 文章多次提及, data summarization methods 会影响用户做出判断以及对后续分析造成影响, 因此需要谨慎选择. 文章归纳了一些 theme, 主要就是从调

Category	Subcategory	Code	Axis — contributing factors	Theme — trends identified from survey data
Data Summarization		Aggregation	Data Summarization	T1 Most summaries use more than one data summarization method.
		Subsampling	Data Summarization	T2 Most summaries use aggregation to summarize data.
		Filtering	Data Summarization × Task	T3 Summaries using aggregation support tasks characterizing the entire dataset.
		Projection	Data Summarization	T4 Aggregation is common across all data types.
Purpose		Exploratory	Data Summarization	T5 Filtering can be used across all data types.
		Confirmatory	Data Summarization × Task	T6 Projection and filtering emphasize similar data characteristics.
		Presentation	Data Summarization	T7 Summaries using subsampling are most common for scientific visualization.
			Data Summarization × Task	T8 Subsampling supports tasks that are statistically robust to random sampling.
Task	Means: Navigation	Browsing	Purpose	T9 Summaries serve as a starting point for analysis.
		Searching	Purpose × Task	T10 Exploratory summaries encode a broad set of data characteristics.
		Elaborating	Purpose	T11 Confirmatory summaries support exploration.
		Summarizing	Purpose × Task	T12 Presentation summaries emphasize a small set of specific characteristics.
	Means: Relation	Comparison	Purpose × Data Summarization	T13 Designs for communicating specific, known information use aggregation.
		Variations	Purpose × Task	T14 Summaries using subsampling emphasize exploration.
		Relation-seeking	Task	T15 Summaries act as roadmaps to guide detailed exploration by interaction.
			Task	T16 Summaries emphasize patterns that characterize all data and dimensions.
	Characteristics: High-level	Trends		
		Outliers		
		Clusters		
		Frequency Distribution Correlation		
Data		Data type Specific data		
Other		Misc. observations		

研的文章提炼出来的一些 insight:

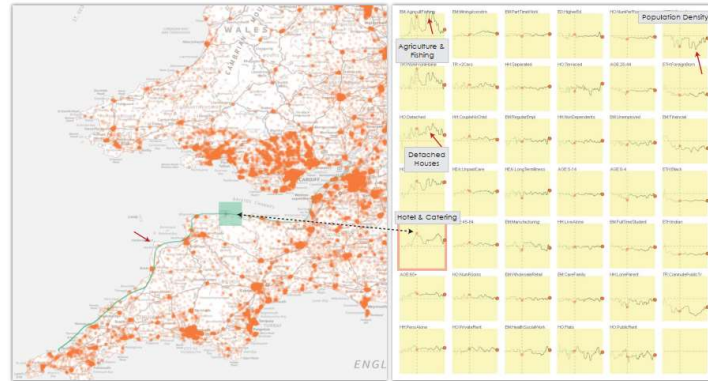
2. VIS18 workshop: VISREG – Visual Summarization and Report Generation: *Developing a formative visual feedback report for data broker*. Data broker 应该是那种数据转卖公司, 他们需要将数据收来, 清洗再卖掉. 那些源头给数据的人往往不按照规范, 有各种各样问题. 本文根据实际处理经验以及用户需求设计了一个 report card. 作为 report 它是静态的; 但用户仍可以有一个交互界面对不同原始数据池进行过滤选择来获取这个 report.

注释: summary 范围应该更广, report 是一个子集, 一般是静态展示用的.

3. CG&A 2016: *Presentation-oriented visualization techniques* 以展示为目的的可视化也将是未来一个热点. 上面的 report 其实也是一种; 而在上面的综述中, presentation 是三种 purpose 之一. 本文分析了此种技术的一些关键点, 他们和传统的可视化或者可视分析系统有啥区别? 以展示为目的的, 需要用户能 get 到 point, 能印象深刻; 可以 specific 而不一定 general; 不一定要非常 scalable, 因为毕竟展示的空间有限. 然后文章列举了一些常见的作品并按照前的关键点加以分析.

4. CG&A 2016: Episogram: Visual Summarization of Egocentric Social Interactions 对于用户交流的一个 visual summarization, 很精致, 但是需要 interaction 来进一步辅助.

5. VIS14/TVCG Attribute Signatures: Dynamic Visual Summaries for Analyzing Multivariate Geographical Data 充分利用了 small multiples 这个方案. 对地理数据探索中不同地点变化\不同范围变化\不同选取粒度变化这三个方面数据变化的查看提出一个统一方案, 来观察随着这些变化而变化的其他值的情况. 如下图, 沿着海岸线, 就是地点变化. 但是尴尬的是, 这三个方面每次只能用一种, 并不能多种一起看.



6. LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records 这是很早以前的 visual report 文章, 针对每个人生成的 (后续有很多作品是以一群人为研究对象的; 但也有 Fan Du 一些文章是以人为作为中心的 event sequence vis, 一般来说这些 event 可观察的状态可枚举).

7 VIS17 Modeling Color Difference for Visualization Design 在 report 生成中, 由于空间限制以及静态的特点(一般会打印出来), 对于各种 encoding 的选择更为高要求. 本文虽不是特别为 report 而写的实验, 但其结论很有用, 不同的 marks, 使得人对于 color difference 的感知程度不一样; 在 report 生成中需要优先考虑这些因素.

8. 另外有综述 Quality Metrics for Information Visualization, A Design Space of visualization tasks.